

CRS Report for Congress

Received through the CRS Web

Adequate Yearly Progress (AYP): Implementation of the No Child Left Behind Act

Updated March 8, 2006

Wayne C. Riddle
Specialist in Education Policy
Domestic Social Policy Division

Adequate Yearly Progress (AYP): Implementation of the No Child Left Behind Act

Summary

Title I, Part A of the Elementary and Secondary Education Act (ESEA), authorizes financial aid to local educational agencies (LEAs) for the education of disadvantaged children and youth at the preschool, elementary, and secondary levels. Over the last several years, the accountability provisions of this program have been increasingly focused on achievement and other outcomes for participating pupils and schools. Since 1994, and particularly under the No Child Left Behind Act of 2001 (NCLBA), a key concept embodied in these requirements is that of “adequate yearly progress (AYP)” for schools, LEAs, and states. AYP is defined primarily on the basis of aggregate scores of various groups of pupils on state assessments of academic achievement. The primary purpose of AYP requirements is to serve as the basis for identifying schools and LEAs where performance is unsatisfactory, so that inadequacies may be addressed first through provision of increased support and, ultimately, a variety of “corrective actions.”

Under the NCLBA, the Title I-A requirements for state-developed standards of AYP were substantially expanded in scope and specificity. Under the NCLBA, AYP calculations must be *disaggregated*—i.e., determined separately and specifically for not only all pupils but also for several demographic groups of pupils within each school, LEA, and state. In addition, while AYP standards had to be applied previously only to pupils, schools, and LEAs participating in Title I-A, AYP standards under the NCLBA must be applied to *all* public schools, LEAs, and to states overall, if a state chooses to receive Title I-A grants. However, corrective actions for failing to meet AYP standards need be applied only to schools and LEAs participating in Title I-A. Another major break with the past is that state AYP standards must now incorporate concrete movement toward meeting an *ultimate goal* of all pupils reaching a proficient or advanced level of achievement within 12 years.

The overall percentage of public schools identified as failing to make AYP for one or more years based on test scores in 2004-2005 was approximately 26% of *all* public schools. The percentage of schools for individual states varied from 2% to 66%. Approximately 14% of Title I-A participating schools were in the “needs improvement” status (i.e., they had failed to meet AYP standards for two or more consecutive years) based on AYP determinations for 2004-2005 and preceding school years.

The AYP provisions of the NCLBA are challenging and complex, and they have generated substantial interest and debate. Debates regarding the NCLBA provisions on AYP have focused on the provision for an ultimate goal, use of confidence intervals and data-averaging, population diversity effects, minimum pupil group size (n), separate focus on specific pupil groups, number of schools identified and state variations therein, the 95% participation rule, state variations in assessments and proficiency standards, and timing. This report will be updated to reflect major new policy developments or available information.

Contents

Background: Title I Outcome Accountability and the AYP Concept	1
General Elements of AYP Provisions	1
Generic AYP Factors	2
AYP Provisions Under the IASA of 1994	5
Concerns About the AYP Provisions of the IASA	6
AYP Under the NCLBA Statute	8
ED Regulations and Guidance on Implementation of the AYP Provisions of the NCLBA	11
Recent Developments	12
Data on Schools and LEAs Identified as Failing to Meet AYP	20
Schools Failing to Meet AYP Standards for <i>One or More</i> Years	21
Schools Failing to Meet AYP Standards for <i>Two or More</i> Consecutive Years	21
LEAs Failing to Meet AYP Standards	23
Issues in State Implementation of the NCLBA Provisions	24
Introduction	24
Ultimate Goal	24
Confidence Intervals and Data-Averaging	27
Population Diversity Effects	28
Minimum Pupil Group Size (n)	28
Separate Focus on Specific Pupil Groups	29
Number of Schools Identified and State Variations Therein	31
95% Participation Rule	33
State Variations in Assessments and Proficiency Standards	33

Adequate Yearly Progress (AYP): Implementation of the No Child Left Behind Act

Background: Title I Outcome Accountability and the AYP Concept

Title I, Part A of the Elementary and Secondary Education Act (ESEA), the largest federal K-12 education program, authorizes financial aid to local educational agencies (LEAs) for the education of disadvantaged children and youth at the preschool, elementary, and secondary levels.

Since the 1988 reauthorization of the ESEA (The Augustus F. Hawkins-Robert T. Stafford Elementary and Secondary School Improvement Amendments of 1988, or “School Improvement Act,” P.L. 100-297), the accountability provisions of this program have been increasingly focused on achievement and other outcomes for participating pupils and schools. Since the subsequent ESEA reauthorization in 1994 (the Improving America’s Schools Act of 1994, P.L. 103-382), and particularly under the No Child Left Behind Act of 2001 (NCLBA, P.L. 107-110), a key concept embodied in these outcome accountability requirements is that of “adequate yearly progress (AYP)” for schools, LEAs, and (more recently) states overall. The primary purpose of AYP requirements is to serve as the basis for identifying schools and LEAs where performance is inadequate, so that these inadequacies may be addressed, first through provision of increased support and, ultimately, through a variety of “corrective actions.”¹

This report is intended to provide an overview of the AYP concept and several related issues, a description of the AYP provisions of the No Child Left Behind Act, and an analysis of the implementation of these provisions by the U.S. Department of Education (ED) and the states. It will be updated when major administrative actions are taken by ED, or substantial new data on state implementation become available.

General Elements of AYP Provisions

ESEA Title I, Part A has included requirements for participating LEAs and states to administer assessments of academic achievement to participating pupils, and to evaluate LEA programs at least every two years, since the program was initiated

¹ These corrective actions, as well as possible performance-based awards, are not discussed in detail in this report. For information on them, see CRS Report RL31487, *Education for the Disadvantaged: Overview of ESEA Title I-A Amendments Under the No Child Left Behind Act*, by Wayne Riddle.

in 1965. However, relatively little attention was paid to school- or LEA-wide outcome accountability until adoption of the School Improvement Act of 1988.² Under the School Improvement Act, requirements for states and LEAs to evaluate the performance of Title I-A schools and individual participating pupils were expanded. In addition, LEAs and states were for the first time required to develop and implement improvement plans for pupils and schools whose performance was not improving. However, in comparison to current Title I-A outcome accountability provisions, these requirements were broad and vague. States and LEAs were given little direction as to how they were to determine whether performance was satisfactory, or how performance was to be defined, with one partial exception.

The exception applied to schools conducting schoolwide programs under Title I-A. In schoolwide programs, Title I-A funds may be used to improve instruction for all pupils in the school, rather than being targeted on only the lowest-achieving individual pupils in the school (as under the other major Title I-A service model, targeted assistance schools). Under the 1988 version of the ESEA, schoolwide programs were limited to schools where 75% or more of the pupils were from low-income families (currently this threshold has been reduced to 40%). The School Improvement Act required schoolwide programs, in order to maintain their special authority, to demonstrate that the academic achievement of pupils in the school was higher than either of the following: (a) the average level of achievement for pupils participating in Title I-A in the LEA overall; or (b) the average level of achievement for disadvantaged pupils enrolled in that school during the three years preceding schoolwide program implementation.

The embodiment of outcome accountability in the specific concept of AYP began with the 1994 Improving America's Schools Act (IASA). Under the IASA, states participating in Title I-A were required to develop AYP standards as a basis for systematically determining whether schools and LEAs receiving Title I-A grants were performing at an acceptable level. Failure to meet the state AYP standards was to become the basis for directing technical assistance, and ultimately corrective actions, toward schools and LEAs where performance was consistently unacceptable.

Generic AYP Factors. Before proceeding to a description of the Title I-A AYP provisions under the IASA of 1994, we outline below the general types of major provisions frequently found in AYP provisions, actual or proposed.

Primary Basis: They are based primarily on aggregate measures of academic achievement by pupils. As long as Title I-A has contained AYP provisions, it has provided that these be based ultimately on *state* standards of curriculum content and pupil performance, and assessments linked to these standards. More specifically, the Title I-A requirements have been focused on the percentage of pupils scoring at the "proficient" or higher level of achievement on state assessments, not a common national standard. However, when AYP provisions were first adopted in 1994, states were given an extended period of time to adopt and implement these standards and

² For additional information on this legislation, see CRS Report 89-7, *Education for Disadvantaged Children: Major Themes in the 1988 Reauthorization of Chapter 1*, by Wayne Riddle (archived, available from author [7-7382] upon request).

assessments, and for a lengthy period after the 1994 amendments, various “transitional” performance standards and assessments were used to measure academic achievement.³

Ultimate Goal: AYP standards may or may not incorporate an ultimate goal, which may be relatively specific and demanding (e.g., all pupils should reach the proficient or higher level of achievement, as defined by each state, in a specified number of years), or more ambiguous and less demanding (e.g., pupil achievement levels must increase in relation to either LEA or state averages or past performance). If there is a specific ultimate goal, there may also be requirements for specific, numerical, annual objectives either for pupils in the aggregate or for each of several pupil groups. The primary purpose of such a goal is to require that levels of achievement continuously increase over time in order to be considered satisfactory.

Subject Areas: With respect to subject areas, AYP standards might focus only on reading and math achievement, or they might include additional subject areas.

Additional Indicators: In addition to pupil scores on assessments, AYP standards often include one or more supplemental indicators, which may or may not be academic. Examples include high school graduation rates, attendance rates, or assessment scores in subjects other than those that are required.

Levels at Which Applied: States may be required to develop AYP standards for, and apply them to, schools, LEAs, and/or for states overall. Further, it may be required that AYP standards be applicable to *all* schools and LEAs, or only to those participating in ESEA Title I-A.

Disaggregation of Pupil Groups: AYP standards might be applied simply to all pupils in a school, LEA, or state, or they might also be applied separately and specifically to a variety of demographic groups of pupils — e.g., economically disadvantaged pupils, pupils with disabilities, pupils in different ethnic or racial groups, or limited English proficient pupils. In a program such as Title I-A, the purpose of which is to improve education for the disadvantaged, it may be especially important to consider selected disadvantaged pupil groups separately, to identify situations where overall pupil achievement may be satisfactory, but the performance of one or more disadvantaged pupil groups is not.

Basic Structure: The basic structure of AYP Models generally falls into one of three general categories. The three basic structural forms for AYP of schools or LEAs are the *group status*, *successive group improvement*, and *individual/cohort growth* models. In the context of these terms, “group” (or “subgroup,” in the case of detailed demographic categories) refers to a collection of pupils that is identified by their grade level and usually other demographic characteristics (e.g., race, ethnicity, or educational disadvantage) as of a point in time. The actual pupils in a “group” may change substantially, or even completely, from one year to the next. In contrast,

³ For additional information on the standard and assessment requirements under ESEA title I-A, see CRS Report RL31407, *Educational Testing: Implementation of ESEA Title I-A Requirements Under the No Child Left Behind Act*, by Wayne C. Riddle.

a “cohort” refers to a collection of pupils in which the *same* pupils are followed from year-to-year.

The key characteristic of the *group status* model is a required threshold level of achievement that is *the same* for all pupil groups, schools, and LEAs statewide in a given subject and grade level. Under this model, performance at a point in time is compared to a benchmark at that time, with no direct consideration of changes over a previous period and whatever the school’s or LEA’s “starting point.” For example, it might be required that 45% or more of the pupils in any of a state’s elementary schools score at the proficient or higher level of achievement in order for a school to make AYP. “Status” models emphasize the importance of meeting certain minimum levels of achievement for all pupil groups, schools, and LEAs, and arguably apply consistent expectations to all.

The key characteristic of the *successive group improvement* model is a focus on the *rate of change* in achievement in a subject area from one year to the next among *groups* of pupils in a grade level at a school or LEA (e.g., the percentage of this year’s 5th grade pupils in a school who are at a proficient or higher level in mathematics compared to the percentage of last year’s 5th grade pupils who were at a proficient or higher level of achievement).

Finally, the key characteristic of the *individual/cohort growth* model is a focus on the *rate of change* over time in the level of achievement among *cohorts* of the same pupils. Growth models are *longitudinal*, based upon the tracking of the same pupils as they progress through their K-12 education careers. While the progress of pupils is tracked individually, results are typically aggregated when used for accountability purposes. Aggregation may be by demographic group, by school or LEA, or other relevant characteristics. In general, growth models would give credit for meeting steps along the way to proficiency in ways that a status model typically does not.

Alternative or “Safe Harbor” Provisions: AYP systems often have alternative provisions under which schools or LEAs that fail to meet the usual requirements may still be deemed to have made AYP if they meet certain specified alternative conditions. For example, under a status model, it might be generally required that 45% or more of the pupils in any of a state’s elementary schools score at the proficient or higher level of achievement in order for the school to make AYP, but a school where aggregate achievement is below this level might still be deemed to have made AYP, through a “safe harbor” provision, if the percentage of pupils at the proficient or higher level in the school is at least 5 percentage points higher than for the previous year. Such a concept may be seen as adding a successive group improvement model element to a status model of AYP.

Assessment Participation Rate: It might be required that a specified minimum percentage of a school’s or LEA’s pupils participate in assessments in order for the school or LEA to be deemed to have met AYP standards. The primary purposes of such a requirement are to assure that assessment results are broadly representative of the achievement level of the school’s pupils, and to minimize the incentives for school staff to discourage test participation by pupils deemed likely to perform poorly on assessments.

Exclusion of Certain Pupils: Beyond general participation rate requirements (see above), states may be specifically required to include, or allowed to exclude, certain groups of pupils in determining whether schools or LEAs meet AYP requirements. For example, statutory provisions might allow the exclusion of pupils who have attended a school for less than one year in determining whether a school meets AYP standards.

Special Provisions for Pupils with Particular Educational Needs: Beyond requirements that all pupils be included in assessments, with accommodations where appropriate, there may be special provisions for limited English proficient (LEP) pupils or pupils with the most significant cognitive disabilities.

Averaging or Other Statistical Manipulation of Data: Finally, there are a variety of ways in which statistical manipulation of AYP-related data or calculations might be either authorized or prohibited. Major possibilities include averaging of test score data over periods of two or more years, rather than use of the latest data in all cases; or the use of “confidence intervals” in calculating whether the aggregate performance of a school’s pupils is at the level specified by the state’s AYP standards. These techniques, and the implications of their use, are discussed further below. In general, their use tends to improve the reliability and validity of AYP determinations, while often reducing the number of schools or LEAs identified as failing to meet AYP standards.

AYP Provisions Under the IASA of 1994

Under the IASA, states were to develop and implement AYP standards soon after enactment. However, states were given several years (generally until the 2000-2001 school year) to develop and implement curriculum content standards, pupil performance standards, and assessments linked to these for at least three grade levels in math and reading.⁴ Thus, during the period between adoption of the IASA in 1994 and of the NCLBA in early 2002, for most states the AYP provisions were based on “transitional” assessments and pupil performance standards that were widely varying in nature. AYP standards based on such “transitional” assessments were considered to be “transitional” themselves, with “final” AYP standards to be based on states’ “final” assessments, when implemented. The subject areas required to be included in state AYP standards (as opposed to required assessments) were not explicitly specified in statute; ED policy guidance required states to include only math and reading achievement in determining AYP. Further, the inclusion in AYP standards of measures other than academic achievement in math and reading on state assessments was optional.

With respect to the ultimate goal of the state AYP standards, the IASA provided broadly that there must be continuous and substantial progress toward a goal of having all pupils meet the proficient and advanced levels of achievement. However,

⁴ For more information on all aspects of the ESEA Title I-A assessment requirements, see CRS Report RL31407, *Educational Testing: Implementation of the ESEA Title I-A Requirements Under the No Child Left Behind Act*, by Wayne C. Riddle.

no timeline was specified for reaching this goal, and most states did not incorporate it into their AYP plans in any concrete way.

The IASA's AYP standards were to be applied to schools and LEAs, but not to the states overall. Further, while states were encouraged to apply the AYP standards to *all* public schools and LEAs, states could choose to apply them only to schools and LEAs participating in Title I-A, and most did so limit their application.

The IASA provided that all relevant pupils⁵ were to be included in assessments and AYP determinations, although assessments were to include results for pupils who had attended a school for less than one year only in tabulating LEA-wide results (i.e., not for individual schools). LEP pupils were to be assessed in the language that would best reflect their knowledge of subjects other than English; and accommodations were to be provided to pupils with disabilities.

Importantly, while the IASA required state assessments to ultimately (by 2000-2001) provide test results that were disaggregated by pupil demographic groups, it did not require such disaggregation of data in AYP standards and calculations. The 1994 statute provided that state AYP standards must consider all pupils, "particularly" economically disadvantaged and LEP pupils, but did not specify that the AYP definition must be based on each of these pupil groups separately. Finally, the statute was silent with respect to data-averaging or other statistical techniques, as well as the basic structure of state AYP standards (i.e., whether a "group status," "successive group improvement," or "individual/cohort growth" model must be employed).

Concerns About the AYP Provisions of the IASA. Thus, the IASA's provisions for state AYP standards broke new ground conceptually, but were comparatively broad and ambiguous. While states were required to adopt and implement at least "transitional" AYP standards, based on "transitional" state assessment results, soon after enactment of the IASA, they were not required to adopt "final" AYP standards, in conjunction with final assessments and pupil performance standards, until the 2000-2001 school year. Further, states were not allowed to implement most corrective actions, such as reconstituting school staff, until they adopted final assessments, so these provisions were not implemented by most states until the IASA was replaced by the NCLBA.

A compilation was prepared by the Consortium for Policy Research in Education (CPRE) of the "transitional" AYP standards that states were applying in administering their Title I-A programs during the 1999-2000 school year.⁶ Overall, according to this compilation, the state AYP definitions for 1999-2000 were widely varied and frequently complex. General patterns in these AYP standards, outlined below, reflect state interpretation of the IASA's statutory requirements.

⁵ All pupils in states where AYP determinations were made for all public schools, or all pupils served by ESEA Title I-A in states where AYP determinations were made only for such schools and pupils.

⁶ See [http://www.cpre.org/Publications/Publications_Accountability.htm].

- Most considered only achievement test scores, but some considered a variety of additional factors, most often dropout rates or attendance rates.
- Often, the state AYP standards set a threshold of some minimum percentage, or minimum rate of increase in the percentage, of pupils at the proficient or higher level of achievement on a composite of state tests. These thresholds were often based, at least in part, on performance of pupils in a school or LEA *relative to* statewide averages or to the school's or the LEA's performance in the previous year. Several states identified schools as failing to make AYP if they fail to meet "expected growth" in performance based on factors such as initial achievement levels and statewide average achievement trends. These thresholds almost never incorporated a "ladder" of movement toward an ultimate goal of all pupils at the proficient level, or otherwise explicitly incorporated an ultimate goal to be met by some specific date.
- While some state AYP standards were based on achievement results for a single year, they were frequently based on two- or three-year rolling averages.
- The AYP standards generally referred only to all pupils in a school or LEA combined, without a specific focus on any pupil demographic groups. However, the AYP standards of some states included a focus on a single category of low-achieving pupils separately from all pupils, and a very few (e.g., Texas) included a specific focus on the performance of several pupil groups (African American, Hispanic, White, or Economically Disadvantaged). One state (New Mexico) compared school scores to predicted scores based on such factors as pupil demographics.
- The state AYP standards under the IASA were sometimes substantially adjusted from year-to-year (often with consequent wide variations in the percentage of Title I-A schools identified as needing improvement). According to CPRE, two states (Iowa and New Hampshire) left AYP standards and determinations almost totally to individual LEAs in 1999-2000.

A report published by ED in 2004, based on state AYP policies for the 2001-2002 school year, contains similar conclusions about state AYP policies in the period immediately preceding implementation of the NCLBA.⁷ There was tremendous variation among the states in the impact of their AYP policies under the IASA on the number and percentage of Title I-A schools and LEAs that were identified as failing to meet the AYP standards. In some states, a substantial majority of Title I-A schools were identified as failing to make AYP, while in others almost no schools were so identified. In July 2002, just before the initial implementation of the new AYP provisions of the NCLBA, ED released a compilation of the number of schools

⁷ U.S. Department of Education. Office of the Undersecretary. Policy and Program Studies Service. *Evaluation of Title I Accountability Systems and School Improvement Efforts (TASSIE): First-Year Findings*. 2004. Hereafter referred to as the TASSIE First-Year Report.

identified as failing to meet AYP standards *for two or more consecutive years* (and therefore identified as being in need of improvement) in 2001-2002 (for most states) or 2000-2001 (in states where 2001-2002 data were not available).⁸ The national total number of these schools was 8,652; the number in individual states ranged from zero in Arkansas and Wyoming to 1,513 in Michigan and 1,009 in California.⁹ While there are obvious differences in the size of these states, there were also wide variations in the percentage of all schools participating in Title I-A that failed to meet AYP for either one year or two or more consecutive years.

AYP Under the NCLBA Statute

The NCLBA provisions regarding AYP may be seen as an evolution of, and to a substantial degree as a reaction to perceived weaknesses in, the AYP requirements of the 1994 IASA. The latter were frequently criticized as being insufficiently specific, detailed, or challenging. Criticism often focused specifically on their failure to focus on specific disadvantaged pupil groups, failure to require continuous improvement toward an ultimate goal, and their required applicability only to schools and LEAs participating in Title I-A, not to all public schools or to states overall.

Under the NCLBA, the Title I-A requirements for state-developed standards of AYP were substantially expanded in scope and specificity. As under the IASA, AYP is defined primarily on the basis of aggregate scores of pupils on state assessments of academic achievement. However, under the NCLBA, state AYP standards must also include at least one additional academic indicator, which in the case of high schools must be the graduation rate. The additional indicators may not be employed in a way that would reduce the number of schools or LEAs identified as failing to meet AYP standards.¹⁰

⁸ See the U.S. Department of Education, "Paige Releases Number of Schools in School Improvement in Each State," press release, July 1, 2002 at [<http://www.ed.gov/news/pressreleases/2002/07/07012002a.html>].

⁹ Another report published by ED in 2004 (the TASSIE First-Year Report — see footnote 7) stated that 8,078 public schools had been identified as failing to meet AYP standards for two or more consecutive years in the 2001-02 school year.

¹⁰ Four bills introduced in the 108th Congress (H.R. 3049, H.R. 4434, H.R. 4464, and S. 2345) and at least two bills introduced in the 109th Congress (H.R. 1506 and S. 724) would specifically authorize states to include measures of *growth* in pupil achievement, on either an individual or cohort basis, as the additional indicator in state AYP standards. Further, five of these bills (all except H.R. 3049) would allow states to use improvement on this additional indicator as a justification for reducing the number of schools identified as failing to meet AYP. In effect, these proposals may be seen as authorizing states to employ a growth-based model of AYP determination as an alternative to the status-based model embodied in the NCLBA statute. Subsequently, as is discussed later in this report and in more detail in a separate report (RL33032, *Adequate Yearly Progress (AYP): Growth Models Under the No Child Left Behind Act*), a growth model pilot project has been announced by ED.

One of the most important differences between AYP standards under the NCLBA and previous requirements is that under the NCLBA, AYP calculations must be *disaggregated* — i.e., they must be determined separately and specifically for not only all pupils but also for several demographic groups of pupils within each school, LEA, and state. Test scores for an individual pupil may be taken into consideration multiple times, depending on the number of designated groups of which they are a member (e.g., a pupil might be considered as part of the LEP and economically disadvantaged groups, as well as the “all pupils” group). The specified demographic groups are as follows:

- economically disadvantaged pupils,
- LEP pupils,
- pupils with disabilities, and
- pupils in major racial and ethnic groups, as well as all pupils.

However, as is discussed further below, there are three major constraints on the consideration of these pupil groups in AYP calculations. First, pupil groups need not be considered in cases where their number is so relatively small that achievement results would not be statistically significant or the identity of individual pupils might be divulged.¹¹ As is discussed further below, the selection of the minimum number (n) of pupils in a group for the group to be considered in AYP determinations has been left largely to state discretion. State policies regarding “n” have varied widely, with important implications for the number of pupil groups actually considered in making AYP determinations for many schools and LEAs, and the number of schools or LEAs potentially identified as failing to make AYP. Second, it has been left to the states to define the “major racial and ethnic groups” on the basis of which AYP must be calculated. And third, as under the IASA, pupils who have not attended the same school for a full year need not be considered in determining AYP for the school, although they are still to be included in LEA and state AYP determinations.

In contrast to the previous statute, under which AYP standards had to be applied only to pupils, schools, and LEAs participating in Title I-A, AYP standards under the NCLBA must be applied to *all* public schools, LEAs, and for the first time to states overall, if a state chooses to receive Title I-A grants. However, corrective actions for failing to meet AYP standards need only be applied to schools and LEAs participating in Title I-A.

Another major break with the past is that state AYP standards must now incorporate concrete movement toward meeting an *ultimate goal* of all pupils reaching a proficient or advanced level of achievement within 12 years (the 2013-2014 school year). The steps — i.e., required levels of achievement — toward meeting this goal must increase in “equal increments” over time. The first increase in the thresholds must occur after no more than two years, and remaining increases at least once every three years. As a result of the initial increase requirement, many states raised their thresholds for the first time in determining AYP based on

¹¹ In addition, program regulations (*Federal Register*, Dec. 2, 2002) do not require graduation rates and other additional academic indicators to be disaggregated in determining whether schools or LEAs meet AYP standards.

assessment scores from the 2004-2005 school year. As is discussed further below, several states have accommodated this requirement in ways that require much more rapid progress in the later years of the period leading up to 2013-2014 than in the earlier period.

The primary basic structure for AYP under the NCLBA is now specified as a group status model. A “uniform bar” approach is employed: states are to set a threshold percentage of pupils at proficient or advanced levels each year that is applicable to all pupil subgroups of sufficient size to be considered in AYP determinations.¹² The threshold levels of achievement are to be set separately for reading and math, and may be set separately for each level of K-12 education (elementary, middle, and high schools). The minimum¹³ *starting point* for the “uniform bar” in the initial period is to be the *greater* of: (a) the percentage of pupils at the proficient or advanced level of achievement for the lowest-achieving pupil group in the base year,¹⁴ *or* (b) the percentage of pupils at the proficient or advanced level of achievement for the lowest-performing quintile (fifth)¹⁵ of schools statewide in the base year.¹⁶ The “uniform bar” must generally be increased at least once every three years, although in the initial period it must be increased after no more than two years.

In determining whether scores for a group of pupils are at the required level, the averaging of scores over two to three years is allowed. In addition, the NCLBA includes a *safe harbor* provision, under which a school that does not meet the standard AYP requirements may still be deemed to meet AYP if it experiences a 10% (not a 10 percentage point) reduction in the gap between 100% and the base year for the specific pupil groups that fail to meet the “uniform bar,” and those pupil groups make progress on at least one other academic indicator included in the state’s AYP standards. As noted earlier, this alternative provision adds successive group improvement as a secondary AYP model under the NCLBA. In addition, as is discussed below, the U.S. Secretary of Education announced in November 2005 a pilot project under which up to 10 states could use a third model of AYP — a “growth model” — for the 2005-2006 school year.

Finally, the NCLBA AYP provisions include an assessment *participation rate* requirement. In order for a school to meet AYP standards, at least 95% of all pupils, as well as at least 95% of each of the demographic groups of pupils considered for

¹² A bill introduced in the 108th Congress, H.R. 3049, would have allowed states to set separate starting points for each relevant pupil group.

¹³ States may, of course, establish starting points above the required level.

¹⁴ The “base year” is the 2001-2002 school year.

¹⁵ This is determined by ranking all public schools (of the relevant grade level) statewide according to their percentage of pupils at the proficient or higher level of achievement (based on all pupils in each school), and setting the threshold at the point where one-fifth of the schools (weighted by enrollment) have been counted, starting with the schools at the lowest level of achievement.

¹⁶ Under program regulations [4 C.F.R. § 200.16(c)(2)], the starting point may vary by grade span (e.g., elementary, middle, etc.) and subject.

AYP determinations for the school or LEA, must participate in the assessments that serve as the primary basis for AYP determinations.¹⁷

ED Regulations and Guidance on Implementation of the AYP Provisions of the NCLBA

States were to begin determining AYP for schools, LEAs, and the states overall based on the NCLBA provisions beginning with the 2002-2003 school year. The deadline for states to submit to ED their AYP standards based on the NCLBA provisions was January 31, 2003, and according to ED, all states met this deadline. On June 10, 2003, ED announced that accountability plans had been approved for all states. However, many of the approved plans required states to take additional actions following submission of their plan.¹⁸

In the period preceding ED's review of state accountability plans under the NCLBA, the Department published two relevant documents. Regulations, published in the *Federal Register* on December 2, 2002, mirrored the detailed provisions in the authorizing statute. The second document, a policy letter published by the Secretary of Education on July 24, 2002,¹⁹ emphasized flexibility, stating that "The purpose of the statute, for both assessments and accountability, is to build on high quality accountability systems that States already have in place, not to require every state to start from scratch." The letter went on to list 10 criteria that it said would be applied by ED in the process of reviewing state AYP standards. These criteria included most, but not all, of the specifications regarding AYP from the authorizing statute and regulations (e.g., applicability to all public schools and their pupils, and specific focus on individual pupil groups). In response to concerns that large numbers of schools might be identified as failing to make AYP (as is discussed further below), ED officials emphasized the importance of taking action to identify and move to improve underperforming schools, no matter how numerous. They also emphasized the possibilities for flexibility and variation in taking corrective actions with respect to schools that fail to meet AYP, depending on the extent to which they fail to meet those standards.

Aspects of state AYP plans that apparently received special attention in ED's reviews included (1) the pace at which proficiency levels are expected to improve (e.g., equal increments of improvement over the entire period, or much more rapid improvement expected in later years than at the beginning); (2) whether schools or LEAs must fail to meet AYP with respect to the *same* pupil group(s), grade level(s) and/or subject areas to be identified as needing improvement, or whether two consecutive years of failure to meet AYP with respect to *any* of these categories

¹⁷ If the number of pupils in a specified demographic group is too small to meet the minimum group size requirements for consideration in AYP determinations, then the participation rate requirement does not apply.

¹⁸ The plans have been posted by ED at [<http://www.ed.gov/admins/lead/account/stateplans03/index.html>].

¹⁹ See [<http://www.ed.gov/news/pressreleases/2002/07/07242002.html>].

should lead to identification;²⁰ (3) the length of time over which pupils should be identified as being LEP; (4) the minimum size of pupil groups in a school in order for the group to be considered in AYP determinations or for reporting of scores; (5) whether to allow schools credit for raising pupil scores from below basic to basic (as well as from basic or below to proficient or above) in making AYP determinations; and (6) whether to allow use of statistical techniques such as “confidence intervals” (i.e., whether scores are below the required level to a statistically significant extent) in AYP determinations.

Recent Developments. On April 7, 2005, the Secretary of Education met with a group of chief state school officers and announced a new, more flexible approach on AYP policies.²¹ This announcement followed a number of instances, beginning in late 2003, in which ED officials published additional regulations and other policy guidance on selected aspects of AYP determination and related assessment issues, in an effort to provide additional flexibility. This guidance has addressed several aspects of AYP implementation that have created particular difficulties for many schools and LEAs: assessment participation rates, calculation of AYP with respect to LEP pupils and pupils with disabilities, plus options for determining AYP in targeted assistance Title I-A programs. While all of these new forms of flexibility may clearly be employed in making future AYP determinations, it is ED’s position that they may not be applied retroactively to AYP determinations for years preceding the time when each particular form of flexibility is announced.²²

At the same time, it has been widely reported that one state has been fined by ED for failure to publish AYP results for schools in a timely manner. According to articles published by these sources in April 2005, the Texas SEA has been fined \$444,282 in Title I-A state administration funds for failure to release AYP results, based on assessment scores for the 2003-2004 school year, until September 30, 2004.²³

Growth Model Pilot Announced by ED. In a press release dated November 18 and in a letter to chief state school officers dated November 21, 2005, the Secretary of Education announced a growth model pilot program under which up to 10 states would be allowed to use growth models to make AYP determinations for

²⁰ ED has approved state accountability plans under which schools or LEAs would be identified as failing to meet AYP only if they failed to meet the required level of performance in the same subject for two or more consecutive years, but has apparently not approved proposals under which a school would be identified only if it failed to meet AYP in the same subject *and* pupil group for two or more consecutive years. A bill was introduced in the 108th Congress, H.R. 3049, that would have allowed states to identify schools as failing to meet AYP standards only if they failed for two or more consecutive years with respect to the same pupil group and subject area.

²¹ See [<http://www.ed.gov/news/pressreleases/2005/04/04072005.html>].

²² This topic is discussed in greater detail in the last section of this report.

²³ See “ED fines Texas for missing AYP reporting deadline,” *Education Daily*, Apr. 26, 2005, p. 1.

the 2005-2006 school year.²⁴ States must apply to ED by February 17, 2006, to obtain permission to participate in the pilot, and the models proposed by the states must meet at least the following criteria (in addition to a variety of criteria applicable to all state AYP policies — e.g., measure achievement separately in reading/language arts and mathematics):

- they must incorporate an ultimate goal of all pupils reaching a proficient or higher level of achievement by the end of the 2013-2014 school year;
- achievement gaps among pupil groups must decline in order for schools or LEAs to meet AYP standards;
- annual achievement goals for pupils must not be set on the basis of pupil background or school characteristics;
- annual achievement goals must be based on performance standards, not past or “typical” performance growth rates;
- the assessment system must produce comparable results from grade to grade and year to year;
- the progress of individual students must be tracked within a state data system.

In addition, applicant states must have their annual assessments for each of grades 3-8 approved by ED, and these assessments must have been in place for at least one year previous to 2005-2006.

States that receive approval to participate in the growth model pilot will have to demonstrate how its impact differs from that of the models specified in the Title I-A statute. This would appear to require such states to calculate AYP using both models for at least the 2005-2006 school year.

The requirements for growth models of AYP under this pilot are relatively restrictive. The models must be consistent with the ultimate goal of all pupils at a proficient or higher level by 2013-2014, a major goal of the statutory AYP provisions of the NCLBA. More significantly, they must incorporate comparable annual assessments, at least for each of grades 3-8 plus at least one senior high school year, and those assessments must be approved by ED and in place for at least one year before 2005-2006. Further, all performance expectations must be individualized, and the state must have an infrastructure of a statewide, longitudinal database for individual pupils. Proposed models would have to be structured around expectations and performance of individual pupils, not demographic groups of pupils in a school or LEA. According to ED, 20 states have submitted applications to be allowed to use growth models to make AYP determinations beginning with either the 2005-2006 or 2006-2007 school years. Currently, these proposals are undergoing peer review, and none has yet been approved.

Pupils with Disabilities. The most substantial of ED’s recent AYP policy changes involves pupils with disabilities. First, regulations addressing the application of the Title I-A standards and assessment requirements to certain pupils

²⁴ See [<http://www.ed.gov/news/pressreleases/2005/11/11182005.html>].

with disabilities were published in the *Federal Register* on December 9, 2003 (pp. 68698-68708). The purpose of these regulations is to clarify the application of standard, assessment, and accountability provisions to pupils “with the most significant cognitive disabilities.” Under the regulations, states and LEAs may adopt alternate assessments *based on alternate achievement standards* — aligned with the state’s academic content standards and reflecting “professional judgment of the highest achievement standards possible” — for a limited percentage of pupils with disabilities.²⁵ The number of pupils whose proficient or higher scores on these alternate assessments may be considered as proficient or above for AYP purposes is limited to a maximum of 1.0% of all tested pupils (approximately 9% of all pupils with disabilities) at the state and LEA level (there is no limit for individual schools). SEAs may request from the U.S. Secretary of Education an exception allowing them to exceed the 1.0% cap statewide, and SEAs may grant such exceptions to LEAs within their state. According to ED staff, three states in 2003-2004 (Montana, Ohio, and Virginia), and four states in 2004-2005 (the preceding three states plus South Dakota), received waivers to go marginally above the 1.0% limit statewide. In the absence of a waiver, the number of pupils scoring at the “proficient or higher” level on alternate assessments, based on alternate achievement standards, in excess of the 1.0% limit is to be added to those scoring “below proficient” in LEA or state-level AYP determinations.

A new ED policy affecting an additional group of pupils with disabilities was announced initially in April 2005, with more details provided on May 10, 2005.²⁶ The new policy is divided into *short-term* (affecting AYP determinations for the 2005-2006 school year, based on the results of assessments administered during the 2004-2005 school year) and *long-term* (affecting subsequent years) *phases*. It is focused on pupils with “persistent academic disabilities,” whose ability to perform academically is assumed to be greater than that of the pupils with “the most significant cognitive disabilities” discussed in the above paragraph, but below that of other pupils with disabilities. In ED’s terminology, these pupils would be assessed using alternate assessments *based on modified achievement standards*. Both the short term and longer term flexibility policies will apply only to states meeting a number of eligibility criteria.

Under the *short-term* policy, affecting AYP determinations for 2005-2006, in eligible states that have not yet adopted modified achievement standards, schools may add to their proficient pupil group a number of pupils with disabilities equal to 2.0% of all pupils assessed (in effect, deeming the scores of all of these pupils to be at the proficient level).²⁷ This policy would be applicable only to schools and LEAs

²⁵ This limitation does *not* apply to the administration of alternate assessments *based on the same standards applicable to all students*, for other pupils with (non-cognitive or less severe cognitive) disabilities.

²⁶ See [<http://www.ed.gov/news/pressreleases/2005/05/05102005.html>].

²⁷ This would be calculated based on statewide demographic data, with the resulting percentage applied to each affected school and LEA in the state. In making the AYP determination using the adjusted data, no further use may be made of confidence intervals or other statistical techniques. (The actual, not just the adjusted, percentage of pupils who
(continued...)

that would otherwise fail meet AYP standards due solely to their pupils with disabilities group. According to ED staff, as of the date of this report, 31 states have been authorized to exercise this short-term flexibility. Alternatively, in eligible states that have adopted modified achievement standards (and applied these to alternate assessments over at least a three-year period), schools and LEAs may count proficient scores for pupils with disabilities on these assessments, subject to a 2.0% (of all assessed pupils) cap at the LEA and state levels.

In order to be eligible for this short-term flexibility, states must: (a) meet the Title I-A and IDEA requirements regarding assessments and AYP determinations for the full range of pupils with disabilities; (b) set a minimum group size (“n”) for pupils with disabilities in AYP determinations equal to that for other pupil groups; (c) “provide information on actions taken to raise achievement for students with disabilities or narrow the achievement gap and evidence that such efforts are improving student achievement,” including aggregate state achievement test scores in reading and mathematics for pupils with disabilities for 2002-2003 and 2003-2004; and (d) submit a variety of commitments and assurances regarding the adoption and implementation of alternate assessments and modified and alternate achievement standards.

The *longer term* policy is embodied in regulations published in the Federal Register on December 15, 2005. These proposed regulations affect standards, assessments, and AYP for a group of pupils with disabilities who are unlikely to achieve grade level proficiency within the current school year, but who are not among those pupils with the most significant cognitive disabilities (whose situation was addressed by an earlier set of regulations, discussed above). For this second group of pupils with disabilities, states would be authorized to develop “modified achievement standards” and alternate assessments linked to these. The modified achievement standards must be aligned with grade-level content standards, but may reflect reduced breadth or depth of grade-level content in comparison to the achievement standards applicable to the majority of pupils. The standards must provide access to grade-level curriculum, and not preclude affected pupils from earning a regular high school diploma.

As with the previous regulations regarding pupils with the most significant cognitive disabilities, there would be no direct limit on the number of pupils who take alternate assessments based on modified achievement standards. However, in AYP determinations, pupil scores of proficient or advanced on alternate assessments based on modified achievement standards may be counted only as long as they do not exceed a number equal to 2.0% of all pupils tested at the state or LEA level (i.e., an estimated 20% of pupils with disabilities); such scores in excess of the limit would be considered “non-proficient.” As with the 1.0% cap for pupils with the most significant cognitive disabilities, this 2.0% cap does not apply to individual schools. In general, LEAs or states could exceed the 2.0% cap only if they did not reach the 1.0% limit with respect to pupils with the most significant cognitive disabilities. Thus, in general, scores of proficient or above on alternate assessments based on

²⁷ (...continued)

are proficient must also be reported to parents and the public.)

alternate *and* modified achievement standards may not exceed a total of 3.0% of all pupils tested at a state or LEA level.²⁸ In particular, states would no longer be allowed to request a waiver of the 1.0% cap regarding pupils with the most significant cognitive disabilities.

The December 15, 2005 proposed regulations also include provisions that are widely applicable to AYP determinations. First, states would no longer be allowed to use varying minimum group sizes (“n”) for different demographic groups of pupils. This would prohibit the frequent practice of setting higher “n” sizes for pupils with disabilities or LEP pupils than for other pupil groups. Second, when pupils take state assessments multiple times, it would no longer be required that only the first test administration be used in AYP determinations; states and LEAs could use the highest score for pupils who take tests more than once. Finally, as with LEP pupils, states and LEAs could include the test scores of former pupils with disabilities in the disability subgroup for up to two years after such pupils have exited special education.²⁹

In summary, if the December 15, 2005 regulations were adopted as proposed, there would be five groups of pupils with disabilities with respect to achievement standards, assessments, and the use of scores in AYP determinations. These groups are summarized below in Table 1.

²⁸ The 3.0% limit might be exceeded for LEAs, but only if — and to the extent that — the SEA waives the 1.0% cap applicable to scores on alternate assessments based on alternate achievement standards.

²⁹ In such cases, the former pupils with disabilities would *not* have to be counted in determining whether the minimum group size was met for the disability subgroup.

Table 1. Categories of Pupils With Disabilities With Respect to Achievement Standards, Assessments, and AYP Determinations Under ESEA Title I-A

Type of achievement standards	Type of assessment	Cap on number of proficient or advanced scores that may be included in AYP determinations (if any)
Regular (i.e., the same as those applicable to pupils generally)	Regular (i.e., the same as that applicable to pupils generally)	None
Regular	Regular with accommodations (e.g., special assistance for those with sight or hearing disabilities)	None
Regular	Alternate assessments based on regular, grade-level achievement standards (e.g., portfolios or performance assessments)	None
Modified academic achievement standards	Assessments based on modified academic achievement standards	In general, 2.0% of all pupils assessed
Alternate achievement standards	Alternate assessments based on alternate achievement standards	In general, 1.0% of all pupils assessed

Participation Rates. On March 29, 2004, ED announced that schools could meet the requirement that 95% or more of pupils (all pupils as well as pupils in each designated demographic group) participate in assessments (in order for the school or LEA to make AYP) on the basis of *average* participation rates for the last two or three years, rather than having to post a 95% or higher participation rate *each* year. In other words, if a particular demographic group of pupils in a public school has a 93% test participation rate in the most recent year, but had a 97% rate the preceding year, the 95% participation rate requirement would be met. In addition, the new guidance would allow schools to exclude pupils who fail to participate in assessments due to a “significant medical emergency” from the participation rate calculations. The new guidance further emphasizes the authority for states to allow pupils who miss a primary assessment date to take make-up tests, and to establish a minimum size for demographic groups of pupils to be considered in making AYP

determinations (including those related to participation rates). According to ED, in some states, as many as 20% of the schools failing to make AYP did so on the basis of assessment participation rates alone. It is not known how many of these schools would meet the new, somewhat more relaxed standard.

LEP Pupils. In a letter dated February 19, and proposed regulations published on June 24, 2004, ED officials announced two new policies with respect to LEP pupils.³⁰ First, with respect to assessments, LEP pupils who have attended schools in the United States (other than Puerto Rico) for less than 10 months must participate in English language proficiency and mathematics tests. However, the participation of such pupils in reading tests (in English), as well as the inclusion of any of these pupils' test scores in AYP calculations, is to be optional (i.e., schools and LEAs need not consider the scores of first year LEP pupils in determining whether schools or LEAs meet AYP standards). Such pupils are still considered in determining whether the 95% test participation has been met.

Second, in AYP determinations, schools and LEAs may continue to include pupils in the LEP demographic category for up to two years after they have attained proficiency in English. However, these formerly LEP pupils need not be included when determining whether a school or LEA's count of LEP pupils meets the state's minimum size threshold for inclusion of the group in AYP calculations, and scores of formerly LEP pupils may not be included in state, LEA, or school report cards. Both these options, if exercised, should increase average test scores for pupils categorized as being part of the LEP group, and reduce the extent to which schools or LEAs fail to meet AYP on the basis of LEP pupil groups.³¹

AYP Determinations for Targeted Assistance Schools. ED has released a February 4, 2004, letter to a state superintendent of education providing more flexibility in AYP determinations for targeted assistance schools.³² Title I-A services are provided at the school level via one of two basic models: targeted assistance schools, where services are focused on individual pupils with the lowest levels of academic achievement, or schoolwide programs, in which Title I-A funds may be used to improve academic instruction for all pupils. Currently, most Title I-A programs are in targeted assistance schools, although the number of schoolwide programs has grown rapidly in recent years, and most pupils served by Title I-A are in schoolwide programs.

This policy letter gives schools and LEAs the option of considering only pupils assisted by Title I-A for purposes of making AYP determinations for individual schools. LEA and state level AYP determinations would still have to be made on the basis of all public school pupils. The impact of this authority, if utilized, is unclear.

³⁰ See *Federal Register*, June 24, 2004, pp. 35462-35465; and [<http://www.ed.gov/nclb/accountability/schools/factsheet-english.html>].

³¹ A bill introduced in the 108th Congress, H.R. 3049, would have authorized the exclusion of scores of LEP pupils who have resided in the United States for less than *three* years, and would allow formerly LEP pupils to be included in that group for AYP calculation purposes indefinitely.

³² See [<http://www.ed.gov/policy/elsec/guid/stateletters/asaypnc.html>].

In schools using this authority, there would be an increased likelihood that pupil demographic groups would be below minimum size to be considered. At the same time, if Title I-A participants are indeed the lowest-performing pupils in targeted assistance schools, it seems unlikely that many schools would choose to base AYP determinations only on those pupils, especially given the current structure of the primary AYP requirements under the NCLBA (i.e., a status model, not a growth model).

Flexibility for Areas Affected By the Gulf Coast Hurricanes.

Following the damage to school systems and dispersion of pupils in the wake of Hurricanes Katrina and Rita in August and September 2005, interest has been expressed by officials of states and LEAs that were damaged by the storms, and/or that enrolled pupils displaced by them, in the possibility of waiving some of the NCLBA's assessment, AYP, or other accountability requirements. In a series of policy letters to chief state school officers (CSSOs), the Secretary of Education has emphasized forms of flexibility already available under current law and announced a number of policy revisions and potential waivers that might be granted in the future.

In a September 29, 2005, letter to all CSSOs,³³ the Secretary of Education noted that they could exercise existing natural disaster provisions of the NCLBA [§1116(b)(7)(D) and (c)(10)(F)] to postpone the implementation of school or LEA improvement designations and corrective actions for schools or LEAs failing to meet AYP standards that are located in the major disaster areas in Louisiana, Alabama, Mississippi, Texas, or Florida, without a specific waiver being required. In addition, waivers of these requirements will be considered for other LEAs or schools heavily affected by enrolling large numbers of evacuee pupils. Further, all affected LEAs and schools could establish a separate subgroup for displaced students in AYP determinations based on assessments administered during the 2005-2006 school year. Pupils would appear only in the evacuee subgroup, not other demographic subgroups (e.g., economically disadvantaged or LEP). Waivers could be requested in 2006 to allow schools or LEAs to meet AYP requirements if only the test scores of the evacuee subgroup would prevent them from making AYP. In any case, all such students must still be assessed and the assessment results reported to the public.

State Revisions of Their Accountability Plans. Over the period following the initial submission and approval of state accountability plans for AYP and related policies in 2003 through the present, many states have proposed a number of revisions to their plans. Sometimes these revisions seem clearly intended to take advantage of new forms of flexibility announced by ED officials, such as those discussed above, while in other cases states appear to be attempting to take advantage of options or forms of flexibility that reportedly been approved for other states previously.

The proposed changes in state accountability plans have apparently almost always been in the direction of increased flexibility for states and LEAs, with reductions anticipated in the number or percentage of schools or LEAs identified as

³³ See [<http://www.ed.gov/policy/elsec/guid/secletter/050929.html>].

failing to make AYP. Issues that have arisen with respect to these changes include a lack of transparency, and possibly inconsistencies (especially over time), in the types of changes that ED officials have approved; debates over whether the net effect of the changes is to make the accountability requirements more reasonable or to undesirably weaken them; concern that the changes may make an already complicated accountability system even more complex; and timing — whether decisions on proposed changes are being made in a timely manner by ED.

The major aspects of state accountability plans for which changes have been proposed *and approved* include the following: (a) changes to take advantage of revised federal regulations and policy guidance regarding assessment of pupils with the most significant cognitive disabilities, LEP pupils, and test participation rates; (b) limiting identification for improvement to schools that fail to meet AYP in the same subject area for two or more consecutive years, and limiting identification of LEAs for improvement to those that failed to meet AYP in the same subject area and across all three grade spans for two or more consecutive years; (c) using alternative methods to determine AYP for schools with very low enrollment; (d) initiating or expanding use of confidence intervals in AYP determinations, including “safe harbor” calculations; (e) changing (usually effectively increasing) minimum group size; and (f) changing graduation rate targets for high schools. Accountability plan changes that have frequently been requested but *not approved* by ED include (a) identification of schools for improvement only if they failed to meet AYP with respect to the same pupil group *and* subject area for two or more consecutive years, and (b) retroactive application of new forms of flexibility to recalculation of AYP for previous years.³⁴

Data on Schools and LEAs Identified as Failing to Meet AYP

Beginning in the summer of 2003, a substantial amount of data has become available on the number of schools and LEAs that failed to meet the AYP standards of the NCLBA for the 2002-2003, 2003-2004, and 2004-2005 school years. A basic problem with almost all such reported data thus far is that they have generally been incomplete (i.e., not all states are included) and subject to change (i.e., the data for several states have been revised one or more times after being initially published, due largely to data corrections and appeals).³⁵ The currently available data reports are discussed below in two categories: reports focusing on the number and percentage of schools failing to meet AYP standards for *one or more years* versus reports on the

³⁴ See Center on Education Policy, *Rule Changes Could Help More Schools Meet Test Score Targets for the No Child Left Behind Act*, Oct. 22, 2004 (available at [<http://www.ctredpol.org>]; Title I Monitor, *Changes in Accountability Plans Dilute Standards, Critics Say*, Nov. 2004; Council of Chief State School Officers, *Revisiting Statewide Educational Accountability Under NCLB*, Sept. 2004, available at [<http://www.ccsso.org>]; and “Requests Win More Leeway Under NCLB,” *Education Week*, July 13, 2005, p. 1.

³⁵ See also “Data Doubts Plague States, Federal Law,” *Education Week*, Jan. 7, 2004.

number and percentage of public identified for school improvement — i.e., they had failed to meet AYP standards for *two, three, four, five, or more consecutive years*.

Schools Failing to Meet AYP Standards for *One or More Years*

Compilations of AYP results for a majority of states for the 2002-2003 through 2004-2005 school years were published in December 2004 and 2005 by *Education Week*.³⁶ While national aggregate comparisons are not possible, due to the number of states for which data were missing for one or more years, these data continue to reflect a pattern of wide variation among states in the percentage of public schools failing to meet AYP standards. Among states providing results, the percentage of public schools failing to meet AYP standards based on assessment results in the 2004-2005 school year ranged from 2% (Wisconsin) to 66% (Hawaii). For 48 states and the District of Columbia, the average share of schools failing to meet AYP standards was 26%. For the 46 states where such a comparison is possible, based on these data, the percentage of public schools failing to make AYP increased between 2003-2004 and 2004-2005 in 24 states, remained the same in two states, and declined in the remaining 20 states. This is largely a reversal of the pattern of change between 2002-2003 and 2003-2004, when among the 36 states where a comparison was possible, the percentage of public schools failing to make AYP increased in only five states, remained the same in one state, and declined in 30 states.

More recently, in February 2006, data on the number of schools failing to meet AYP standards based on assessment results for the 2004-2005 school year were published in *Education Daily*.³⁷ Based on data collected from all states except Arkansas, it was reported that 22,868 schools, constituting 25.6% of all public schools, failed to meet AYP standards for 2004-2005. Other reported results were similar to those described in the preceding paragraph.

Schools Failing to Meet AYP Standards for *Two or More Consecutive Years*

The most recent and complete published data on schools identified for improvement appears in the February 2006 *Education Daily* article discussed above. According to this survey, 11,524 schools, or 12.9% of all public schools, were identified as needing improvement based on assessment results for the 2004-2005 and preceding school years. In several states, this group included at least some non-Title I schools. With respect to the various stages of school improvement, a total of 3,757 schools had failed to meet AYP standards for two consecutive years, 3,696 for three consecutive years, 1,254 for four consecutive years, and 1,847 for five or more consecutive years.

³⁶ See “Taking Root,” *Education Week*, Dec. 8, 2004, p. 1, and “Room to Maneuver,” *Education Week*, Dec. 14, 2005, p. S1.

³⁷ “Data analysis finds more schools subject to sanctions,” *Education Daily*, Feb. 16, 2006, pp. 1-2.

Earlier, in December 2005, a survey of the number of schools identified for improvement — i.e., had failed to meet AYP standards for two or more consecutive years on the basis of assessment results for 2004-05 and preceding school years — was published in *Education Week*.³⁸ The survey included all states except one (Nebraska) and found that on average, 14% of public schools had been identified as needing improvement. Again, the proportion varied widely among the states, ranging from 1% in Kansas and Utah to 48% in Hawaii.

Finally, in March 2005, the non-governmental Center on Education Policy published its report, “From the Capital to the Classroom: Year 3 of the No Child Left Behind Act.”³⁹ This report provides estimates, based on a nationally representative sample survey of LEAs, of the number of Title I-A schools identified for improvement — i.e., they had failed to meet AYP standards for two or more consecutive years — for each of the 2002-2003, 2003-2004, and 2004-2005 school years (in each case, based on assessment scores for the preceding two or more school years). According to this report, the percentage of Title I-A schools identified for improvement has remained relatively constant, at 13%-14% per year, or approximately 6,000 schools. According to the authors of this report, schools in very large, urban LEAs, and middle schools in general, are most likely to be identified for improvement, while schools in small, rural LEAs, and elementary schools in general, are least likely to be so identified. In addition to these schools participating in Title I-A, an estimated 2,400 non-Title I-A schools were identified for improvement for 2004-2005.⁴⁰

³⁸ See “Room to Maneuver,” *Education Week*, Dec. 14, 2005, p. S1.

³⁹ Available at [<http://www.ctredpol.org/pubs/nclby3/>].

⁴⁰ Also in Mar. 2005, a report was published by the Education Commission of the States (ECS) that provided a great deal of state-by-state data on schools identified for improvement. However, the ECS has since stopped disseminating the report and removed it from its Internet site, stating that the organization “regrets that news reports in several states misinterpreted the data in our recent report about the number of schools not meeting standards of Adequate Yearly Progress (AYP) under the No Child Left Behind Act. No conclusion about the quality of individual schools or a state’s education system can be drawn simply by considering the status of a school or group of schools on the AYP standard. In view of the misinterpretations that can be drawn from comparison of state AYP data, ECS has removed the publication, “Schools’ Status in School Improvement Categories,” from its website and from publication. ECS apologizes for the unwarranted negative assessments of any state’s schools that resulted from this document.”

The report, “Schools’ Status in School Improvement Categories,” had provided data for each of the 50 states (but not the District of Columbia and Puerto Rico) on the number of schools identified for improvement for the 2004-2005 school year because they had failed to meet state AYP standards *for two or more consecutive school years*. It also provided detailed information on the number of schools facing the varying stages of school improvement specified under the NCLBA — i.e., the number of schools that had failed to meet AYP standards for two (only) versus three, four, or five or more consecutive years.

According to the report, only 14 states had any schools in the final stage of school improvement, restructuring, during the 2004-2005 school year. Most of these schools were
(continued...)

LEAs Failing to Meet AYP Standards

As mentioned above, states receiving ESEA Title I-A grants are required to establish and implement AYP standards not only for all public schools in the state, but also for LEAs overall, and the state as a whole. While most attention, in both the statute and implementation activities, thus far has been focused on application of the AYP concept to schools, a limited amount of information is becoming available about LEAs that fail to meet AYP requirements, and the consequences for them. According to the *Education Daily* survey referred to above (in the discussion of schools failing to meet AYP standards), 3,281 LEAs, or 23.7% of all LEAs, failed to meet AYP standards on the basis of assessment results for the 2004-2005 school year. Of these, 1,712 LEAs (12.4% of all LEAs) were identified for improvement as a result of failing to meet AYP standards for two or more consecutive years.

Earlier, the Year 3 report of the Center on Education Policy on No Child Left Behind implementation⁴¹ found that an estimated 10% of all LEAs were identified as needing improvement, based on assessment results for the 2003-2004 and immediately preceding school years.⁴² According to this report, the odds of being identified for improvement were much greater for urban (25% identified) or very large (52%) LEAs than for rural (7%) or small (also 7%) LEAs. According to the authors of the CEP report, states “are just starting to face the prospect of administering corrective action programs for districts,”⁴³ and only very limited information is available on actions, if any, taken with respect to the LEAs identified for improvement.

A large number of states have recently adopted policies under which LEAs would be identified as needing improvement only if they failed to make AYP in the same subject (reading or mathematics) in each of three grade levels (elementary, middle, and high) for two or more consecutive school years. According to a recent

⁴⁰ (...continued)

reported to be in Alabama, Georgia, Hawaii, Maryland, Michigan, New York, and Pennsylvania. These would be schools that had already failed to meet AYP standards for multiple consecutive years before enactment of the NCLBA, and had consistently failed to meet the revised AYP standards afterward. Overall, it was reported that during the 2004-2005 school year, the percentage of all public schools that were in some stage of school improvement varied widely among the 50 states — from 1% or below in states such as Iowa, Kansas, or Nebraska, to 28% in Florida, 37% in Georgia, and 47% in Hawaii. The nationwide average was reported to be 11% of public schools at some stage of school improvement.

⁴¹ Available at [<http://www.ctredpol.org/pubs/nclby3/>].

⁴² While there were AYP requirements for LEAs under the IASA, the application of these requirements by states was apparently quite uneven, and the provisions for consequences for LEAs that failed to meet AYP standards for multiple years were minimal.

⁴³ Center on Education Policy, “From the Capital to the Classroom, Year 3 of the No Child Left Behind Act,” March 2005, p. 14. See also: Center on Education Policy, “Identifying School Districts for Improvement and Corrective Action,” March 2005. Both reports are available at [<http://www.ctredpol.org/>].

study of NCLBA implementation in six states by the Harvard Civil Rights Project, this has substantially increased the proportion of LEAs identified for improvement that serve central city areas and racially diverse and/or high poverty pupil populations.⁴⁴

Issues in State Implementation of the NCLBA Provisions

Introduction

The primary challenge associated with the AYP concept is to develop and implement school, LEA, and state performance measures that are: (a) challenging, (b) provide meaningful incentives to work toward continuous improvement, (c) are at least minimally consistent across LEAs and states, and (d) focus attention especially on disadvantaged pupil groups. At the same time, it is generally deemed desirable that AYP standards should allow flexibility to accommodate myriad variations in state and local conditions, demographics, and policies, and avoid the identification of so many schools and LEAs as failing to meet the standards that morale declines significantly systemwide and it becomes extremely difficult to target technical assistance and corrective actions on low-performing schools. The AYP provisions of the NCLBA are challenging and complex, and have generated substantial criticism from several states, LEAs, and interest groups. Many critics are especially concerned that efforts to direct resources and apply corrective actions to low-performing schools would likely be ineffective if resources and attention are dispersed among a relatively large proportion of public schools. Others defend the NCLBA's requirements as being a measured response to the weaknesses of the pre-NCLBA AYP provisions, which were much more flexible but, as discussed above, had several weaknesses.

The remainder of this report provides a discussion and analysis of several specific aspects of the NCLBA's AYP provisions that have attracted significant attention and debate. These include the provision for an ultimate goal, use of confidence intervals and data-averaging, population diversity effects, minimum pupil group size (n), separate focus on specific pupil groups, number of schools identified and state variations therein, the 95% participation rule, state variations in assessments and proficiency standards, and timing.

It should be noted that this report focuses on issues that have arisen in the implementation of the NCLBA provisions on AYP. As such, it generally does not focus on *alternatives* to the current statutory provisions of the NCLBA. A forthcoming CRS report will discuss a major possible alternative to the currently approved AYP provisions of the NCLBA (individual/cohort growth models of AYP).

⁴⁴ Harvard Civil Rights Project, "Changing NCLB Accountability Standards: Implications for Racial Equity," June 2005, available at [<http://www.civilrightsproject.harvard.edu>].

Ultimate Goal

The required incorporation of an ultimate goal — of all pupils at a proficient or higher level of achievement within 12 years of enactment — is one of the most significant differences between the AYP provisions of the NCLBA and those under previous legislation. Setting such a date is perhaps the primary mechanism requiring state AYP standards to incorporate annual increases in expected achievement levels, as opposed to the relatively static expectations embodied in most state AYP standards under the previous IASA. Without an ultimate goal of having all pupils reach the proficient level of achievement by a specific date, states might simply establish relative goals (e.g., performance must be as high as the state average) that provide no real movement toward, or incentives for, significant improvement, especially among disadvantaged pupil groups.

Nevertheless, a goal of having all pupils at a proficient or higher level of achievement, within 12 years or any other specified period of time, may be easily criticized as being “unrealistic,” if one assumes that “proficiency” has been established at a challenging level. Proponents of such a demanding ultimate goal argue that schools and LEAs frequently meet the goals established for them, even rather challenging goals, *if* the goals are very clearly identified, defined, and established, if they are attainable, and if it is made visibly clear that they will be expected to meet them. This is in contrast to a pre-NCLBA system under which performance goals were often vague, undemanding, and poorly communicated, with few, if any, consequences for failing to meet them. A demanding goal might maximize efforts toward improvement by state public school systems, *even if* the goal is not met. Further, if a less ambitious goal were to be adopted, what lower level of pupil performance might be acceptable, and for which pupils?

At the same time, by setting deadlines by which all pupils must achieve at the proficient or higher level, the AYP provisions of the NCLBA create an incentive for states to weaken their pupil performance standards to make them easier to meet. In many states, only a minority of pupils (sometimes a small minority) are currently achieving at the proficient or higher level on state reading and mathematics assessments. Even in states where the percentage of all pupils scoring at the proficient or higher level is substantially higher, the percentage of those in many of the pupil groups identified under the NCLBA’s AYP provisions is substantially lower.⁴⁵ It would be extremely difficult for such states to reach a goal of 100% of their pupils at the proficient level, even within 10-12 years, without reducing their performance standards.

There has thus far been some apparent movement toward lowering proficiency standards in a small number of states. Reportedly, a few states have redesignated lower standards (e.g., “basic” or “partially proficient”) as constituting a “proficient” level of performance for Title I-A purposes, or established new “proficient” levels

⁴⁵ For example, while 65%-74% of *all* pupils scored at the proficient or higher level on reading and math tests administered to pupils in Indiana in grades 3, 6, and 8, in September 2003, the percentage of Indiana pupils with disabilities who scored at the proficient or higher level on these tests ranged from 20% to 47%.

of performance that are below levels previously understood to constitute that level of performance, and other states have considered such actions.⁴⁶ For example, in submitting its accountability plan (which was approved by ED), Colorado stated that it would deem students performing at both its “proficient” and “partially proficient” levels, as defined by that state, as being “proficient” for NCLBA purposes.⁴⁷ In its submission, the state argued that “Colorado’s standards for all students remain high in comparison to most states. Colorado’s basic proficiency level on CSAP is also high in comparison to most states.” Similarly, Louisiana decided to identify its “basic” level of achievement as the “proficient” level for NCLBA purposes, stating that “[t]hese standards have been shown to be high; for example, equipercentile equating of the standards has shown that Louisiana’s ‘Basic’ is somewhat more rigorous than NAEP’s ‘Basic.’ In addition, representatives from Louisiana’s business community and higher education have validated the use of ‘Basic’ as the state’s proficiency goal.”⁴⁸

This is an aspect of the NCLBA’s AYP provisions on which there will likely be continuing debate and, possibly, future adjustments. It is unlikely that any state, or even any school or LEA of substantial size and a heterogeneous pupil population, will meet the NCLBA’s ultimate AYP goal, unless state standards of proficient performance are significantly lowered and/or states aggressively pursue the use of such statistical techniques as setting high minimum group sizes and confidence intervals (described below) to substantially reduce the range of pupil groups considered in AYP determinations and/or effectively lower required achievement level thresholds.

Some states have addressed this situation, at least in the short run, by “backloading” their AYP standards, requiring much more rapid improvements in performance at the end of the 12-year period than at the beginning. These states have followed the letter of the statutory language that requires increases of “equal increments” in levels of performance after the first two years, and at least once every three years thereafter.⁴⁹ However, they have “backloaded” this process by, for example, requiring increases only once *every two-three years* at the beginning, then requiring increases of the same degree *every year* for the final years of the period leading up to 2013-2014. For example, both Indiana and Ohio established incremental increases in the threshold level of performance for schools and LEAs that are equal in size, and that are to take effect in the school years beginning in 2004, 2007, 2010, 2011, 2012, and 2013. As a result, the required increases *per year* are

⁴⁶ See, for example, “States Revise the Meaning of ‘Proficient’,” *Education Week*, Oct. 9, 2002.

⁴⁷ See [<http://www.ed.gov/admins/lead/account/stateplans03/cocsa.pdf>], p. 7.

⁴⁸ See [<http://www.ed.gov/admins/lead/account/stateplans03/lacsa.doc>], p 12.

⁴⁹ According to Section 1111(b)(2)(H), “Each State shall establish intermediate goals for meeting the requirements, ..., of this paragraph and that shall — (i) increase in equal increments over the period covered by the State’s timeline....” The program regulations also would seem to require increases in equal increments: “Each State must establish intermediate goals that increase in equal increments over the period covered by the timeline....” (34 C.F.R. § 200.17).

three times greater during 2010-2013 than in the 2004-2009 period. These states may be trying to postpone required increases in performance levels until the NCLBA provisions are reconsidered, and possibly revised, by a future Congress.

Confidence Intervals and Data-Averaging

Many states have used one or both of a pair of statistical techniques to attempt to improve the validity and reliability of AYP determinations. Use of these techniques also tends to have an effect, whether intentional or not, of reducing the number of schools or LEAs identified as failing to meet AYP standards.

The averaging of test score results for various pupil groups over two- or three-year periods is explicitly authorized under the NCLBA, and this authority is used by many states. In some cases, schools or LEAs are allowed to select whether to average test score data, and for what period (two years or three), whichever is most favorable for them. As discussed above, recent policy guidance also explicitly allows the use of averaging for participation rates. A bill introduced in the 108th Congress, H.R. 3049, would have authorized expanded forms of data-averaging by states in AYP calculations.

The use of another statistical technique was not explicitly envisioned in the drafting of the NCLBA's AYP provisions, but its inclusion in the accountability plans of several states has been approved by ED. This is the use of "confidence intervals," usually with respect to test scores, but in a couple of states also to the determination of minimum group size (see below). This concept is based on the assumption that any test administration represents a "sample survey" of pupils' educational achievement level. As with all sample surveys, there is a degree of uncertainty regarding how well the sample results — average test scores for the pupil group — reflect pupils' actual level of achievement. As with surveys, the larger the number of pupils in the group being tested, the greater the probability that the group's average test score will represent their true level of achievement, all else being equal. Put another way, confidence intervals are used to evaluate whether achievement scores are below the required threshold *to a statistically significant extent*.

"Confidence intervals" may be seen as "windows" surrounding a threshold test score level (i.e., the percentage of pupils at the proficient or higher level required under the state's AYP standards).⁵⁰ The size of the window varies with respect to the number of pupils in the relevant group who are tested, and with the desired degree of probability that the group's average score represents their true level of achievement. This is analogous to the "margin of error" commonly reported along with opinion polls. While test results are not based on a small sample of the relevant population, as are opinion poll results, since the tests are to be administered to the full "universe" of pupils, the results from any particular test administration are considered to be only estimates of pupils' true level of achievement, and thus the "margin of error" or "confidence interval" concepts are deemed relevant to all test

⁵⁰ Alternatively, the confidence interval "window" may be applied to average test scores for each relevant pupil group, that would be compared to a fixed threshold score level to determine whether AYP has been met.

scores. The probability, or level of confidence, is typically set at 95%, but in some cases may be 99% — i.e., it is 95% (or 99%) certain that the true achievement level for a group of pupils is within the relevant confidence interval of test scores above and below the average score for the group. All other relevant factors being equal, the smaller the pupil group, and the higher the desired degree of probability, the larger is the window surrounding the threshold percentage.

For example, consider a situation where the threshold percentage of pupils at the proficient or higher level of achievement in reading for elementary schools required under a state’s AYP standards is 40%. Without applying confidence intervals, a school would simply fail to make AYP if the average scores of all of its pupils, or of any of its relevant pupil groups meeting minimum size thresholds, is below 40%. In contrast, if confidence intervals are applied, windows are established above and below the 40% threshold, turning the threshold from a single point to a variable range of scores. The size of this score range or window will vary depending on the size of the pupil group whose average scores are being considered, and the desired degree of probability (95% or 99%) that the average achievement levels for pupils in each group are being correctly categorized as being “truly” below the required threshold. In this case, a school would fail to make AYP with respect to a pupil group only if the average score for the group is below the lowest score in that range.

The use of confidence intervals to determine whether group test scores fall below required thresholds to a statistically significant degree improves the validity of AYP determinations, and addresses the fact that test scores for any group of pupils will vary from one test administration to another, and these variations may be especially large for a relatively small group of pupils. At the same time, the use of confidence intervals reduces the likelihood that schools or (to a lesser extent) LEAs will be identified as failing to make AYP. Also, for relatively small pupil groups and high levels of desired accuracy (especially a 99% probability), the size of confidence intervals may be rather large. Ultimately, the use of this technique may mean that the average achievement levels of pupil groups in many schools will be well below 100% proficiency by 2013-2014, yet the schools would still meet AYP standards because the groups’ scores are within the relevant confidence interval.

Population Diversity Effects

Minimum Pupil Group Size (n). Another important technical factor in state AYP standards is the establishment of the minimum size (n) for pupil groups to be considered in AYP calculations. The NCLBA recognizes that in the disaggregation of pupil data for schools and LEAs, there might be pupil groups that are so small that average test scores would not be statistically reliable, or the dissemination of average scores for the group might risk violation of pupils’ privacy rights.

Both the statute and ED regulations and other policy guidance have left the selection of this minimum number to state discretion. While most states have reportedly selected a minimum group size between 30 and 50 pupils, the range of selected values for “n” is rather large, varying from as few as five to as many as 200

pupils⁵¹ under certain circumstances. Two sparsely populated states (Montana and North Dakota) set no specific level for “n,” relying only on the use of confidence intervals (see above) to establish reliability of test results. While most states have set a standard minimum size for all pupil groups, some states have established higher levels of “n” for pupils with disabilities.⁵²

In general, the higher the minimum group size, the less likely that many pupil groups will actually be separately considered in AYP determinations. (Pupils will still be considered, but only as part of the “all pupils” group, or possibly other specified groups.) This gives schools and LEAs fewer thresholds to meet, and reduces the likelihood that they will be found to have failed to meet AYP standards. At the same time, relatively high levels for “n” weaken the NCLBA’s specific focus on a variety of pupil groups, many of them disadvantaged, such as LEP pupils, pupils with disabilities, or economically disadvantaged pupils.

Separate Focus on Specific Pupil Groups. There are several ongoing issues regarding the NCLBA’s requirement for disaggregation of pupil achievement results in AYP standards — i.e., the requirement that a variety of pupil groups be separately considered in AYP calculations. The first of these was discussed immediately above — the establishment of minimum group size, with the possible result that relatively small pupil groups will not be considered in the schools and LEAs of states that set “n” at a comparatively high level, especially in states that set a higher level for certain groups (e.g., pupils with disabilities) than others.

A second issue arises from the fact that the definition of the specified pupil groups has been left essentially to state discretion. This is noteworthy particularly with respect to two groups of pupils — LEP pupils and pupils in major racial and ethnic groups. Regarding LEP pupils, many have been concerned about the difficulty of demonstrating that these pupils are performing at a proficient level if this pupil group is defined narrowly to include only pupils unable to perform in regular English-language classroom settings. In other words, if pupils who no longer need special language services are no longer identified as being LEP, how will it be possible to bring those who are identified as LEP up to a proficient level of achievement?

In developing their AYP standards, some states addressed this concern by including pupils in the LEP category for one or more years after they no longer need special language services. As was discussed above, ED has recently published policy guidance encouraging all states to follow this approach, allowing them to continue to include pupils in the LEP group for up to two years after being mainstreamed into regular English language instruction, and further allowing the scores of LEP pupils

⁵¹ In Texas, the minimum group size for pupil groups (other than the “all pupils” group) is 50 if that group constitutes at least 10% of the pupils in a school or LEA, and up to 200 if below 10% of the enrollment. In California, the minimum group size is 100 in general, although it is 50 if such students constitute 15% or more of enrollment.

⁵² A higher minimum group size for pupils with disabilities is allowed under current ED policy guidance, and would have been explicitly authorized under a bill introduced in the 108th Congress: H.R. 3049.

to be excluded from AYP calculations for the first year of pupils' enrollment in United States schools. If widely adopted, these policies should reduce the extent that schools or LEAs are identified as failing to meet AYP standards on the basis of the LEP pupil group.

Another aspect of this issue arises from the discretion given to states in defining "major racial and ethnic groups." Neither the statute nor ED has defined this term. Some states defined the term relatively comprehensively (e.g., Maryland includes American Indian, African American, Asian, white, and Hispanic pupil groups) and some more narrowly (e.g., Texas identifies only three groups — white, African American, and Hispanic). A more narrow interpretation may reduce the attention focused on excluded pupil groups. It would also reduce the number of different thresholds some schools and LEAs would have to meet in order to make AYP.

A final, overarching issue arises from the relationship between pupil diversity in schools and LEAs and the likelihood of being identified as failing to meet AYP standards. All other relevant factors being equal (especially the minimum group size criteria), the more diverse the pupil population, the more thresholds a school or LEA must meet in order to make AYP. While in a sense this was an intended result of legislation designed to focus (within limits) on all pupil groups, the impact of making it more difficult for schools and LEAs serving diverse populations to meet AYP standards may also be seen as an unintended consequence of the NCLBA. This issue has been analyzed in a recent study by Thomas J. Kane and Douglas O. Staiger, who concluded that such "subgroup targets cause large numbers of schools to fail... arbitrarily single out schools with large minority subgroups for sanctions... or statistically disadvantage diverse schools that are likely to be attended by minority students.... Moreover, while the costs of the subgroup targets are clear, the benefits are not. Although these targets are meant to encourage schools to focus more on the achievement of minority youth, we find no association between the application of subgroup targets and test score performance among minority youth."⁵³ Thus far, insufficient data are available to evaluate whether this prediction is being borne out in practice.

An additional study published by Policy Analysis for California Education (PACE)⁵⁴ found that when comparing public schools in California with similar aggregate pupil achievement levels, schools with larger numbers of different NCLBA-relevant demographic groups were substantially less likely to have met AYP standards in the 2002-2003 school year. Similarly when comparing California public schools with comparable percentages of pupils from low-income families, schools with larger numbers of relevant demographic groups of pupils were much less likely to have met AYP.

⁵³ Thomas J. Kane and Douglas O. Staiger, "Unintended Consequences of Racial Subgroup Rules," in Paul Peterson and Martin West, eds., *No Child Left Behind? The Politics and Practice of School Accountability* (Washington: Brookings Institution Press, 2003), pp. 152-176.

⁵⁴ John R. Novak and Bruce Fuller, *Penalizing Diverse Schools?* PACE Policy Brief 03-4, Dec. 2003.

However, without specific requirements for achievement gains by each of the major pupil groups, it is possible that insufficient attention would be paid to the performance of the disadvantaged pupil groups among whom improvements are most needed, and for whose benefit the Title I-A program was established. Under previous law, without an explicit, specific requirement that AYP standards focus on these disadvantaged pupil groups, most state AYP definitions considered only the performance of all pupils combined. And it is theoretically possible for many schools and LEAs to demonstrate substantial improvements in achievement by their pupils overall while the achievement of their disadvantaged pupils does not improve significantly, at least until the ultimate goal of all pupils at the proficient or higher level of achievement is approached. This is especially true under a “status” model of AYP such as the one in the NCLBA, under which advantaged pupil groups may have achievement levels well above what is required, and an overall achievement level could easily mask achievement well below the required threshold by various groups of disadvantaged pupils.

Number of Schools Identified and State Variations Therein

As was discussed earlier, concern has been expressed by some analysts since early debates on the NCLBA that a relatively high proportion of schools would fail to meet AYP standards. While the numbers of schools failing to meet AYP on the basis of assessment results in 2004-2005 remain somewhat in flux, it would appear that approximately 26% of all public schools nationwide fell into this category, and that approximately 14% of all public schools were identified as needing improvement (i.e., failed to meet AYP standards for two or more consecutive years) on the basis of AYP determinations for 2004-2005 and the immediately preceding school years. Future increases in performance thresholds, as the ultimate goal of all pupils at the proficient or higher level of achievement is approached, as well as implementation of tests in additional grades in many states,⁵⁵ may result in higher percentages of schools failing to make AYP.

In response to these concerns, ED officials have emphasized the importance of taking action to identify and move to improve underperforming schools, no matter how numerous. They have also emphasized the possibilities for flexibility and variation in taking corrective actions with respect to schools that fail to meet AYP, depending on the extent to which they fail to meet those standards. It should also be re-emphasized that many of the schools reported as having failed to meet AYP standards have failed to meet AYP for one year only, while the NCLBA requires that a series of actions be taken only with respect to schools or LEAs participating in ESEA Title I-A that fail to meet AYP for *two or more consecutive years*.

⁵⁵ Several states do not currently administer standards-based assessments in mathematics and reading in each of grades 3-8, or assessments in science at three grade levels, as is required in future years under the NCLBA. As such assessments are administered to pupils in additional grades and subject areas, there will be increases in the number of pupil groups meeting minimum size thresholds to be considered in AYP determinations, and possibly also increases in the number of different test score thresholds that many schools and LEAs have to meet.

Further, some analysts argue that a set of AYP standards that one-quarter or more of public schools fail to meet may accurately reflect pervasive weaknesses in public school systems, especially with respect to the performance of disadvantaged pupil groups. To these analysts, the identification of large percentages of schools is a positive sign of the rigor and challenge embodied in the NCLBA's AYP requirements, and is likely to provide needed motivation for significant improvement (and ultimately a reduction in the percentage of schools so identified).

Others have consistently expressed concern about the accuracy and efficacy of an accountability system under which such a high percentage of schools is identified as failing to make adequate progress, with consequent strain on financial and other resources necessary to provide technical assistance, public school choice and supplemental services options, as well as other corrective actions. In addition, some have expressed concern that schools might be more likely to fail to meet AYP simply because they have diverse enrollments, and therefore more groups of pupils to be separately considered in determining whether the school meets AYP standards. They also argue that the application of technical assistance and, ultimately, corrective actions to such a high percentage of schools will dilute available resources to such a degree that these responses to inadequate performance would be insufficient to markedly improve performance. A few analysts even speculate that the AYP system under the NCLBA is intended to portray large segments of American public education as having "failed," leading to proposals for large scale privatization of elementary and secondary education.⁵⁶

The proportion of public schools identified as failing to meet AYP standards is not only relatively large in the aggregate, but also varies widely among the states. As was discussed above, the percentage of public schools identified as failing to make AYP on the basis of assessment results for 2004-2005 ranged from 2% to approximately two-thirds for the states for which data are available. This result is somewhat ironic, given that one of the major criticisms of the pre-NCLBA provisions for AYP was that they resulted in a similarly wide degree of state variation in the proportion of schools identified, and the more consistent structure required under the NCLBA was widely expected to lead to at least somewhat greater consistency among states in the proportion of schools identified.

It seems likely that the pre-NCLBA variations in the proportion of schools failing to meet AYP reflected large differences in the nature and structure of state AYP standards, as well as major differences in the nature and rigor of state pupil performance standards and assessments. While the basic structure of AYP definitions is now substantially more consistent across states, significant variations remain with respect to the factors discussed in this section of the report (such as minimum group size), and substantial differences in the degree of challenge embodied in state standards and assessments remain. Overall, it seems likely that the key influences determining the percentage of a state's schools that fails to make AYP include (in no particular order): (1) degree of rigor in state content and pupil performance standards; (2) minimum pupil group size (n) in AYP determinations; (3)

⁵⁶ See Alfie Kohn, "Test Today, Privatize Tomorrow: Using Accountability to 'Reform' Public Schools to Death," *Phi Delta Kappan*, vol. 85, no. 8 (Apr. 2004), pp. 568-577.

use of confidence intervals in AYP determinations (and whether at a 95% or 99% level of confidence); (4) extent of diversity in pupil population; (5) extent of communication about, and understanding of, the 95% test participation rule, and (6) possible actual differences in educational quality.

95% Participation Rule

It appears that in many cases, schools or LEAs have failed to meet AYP solely because of low participation rates in assessments — i.e., fewer than 95% of all pupils, or of pupils in relevant demographic groups meeting the minimum size threshold, took the assessments. While, as discussed above, ED recently published policy guidance that relaxes the participation rate requirement somewhat — allowing use of average rates over two- to three-year periods, and excusing certain pupils for medical reasons — the high rate of assessment participation that is required in order for schools or LEAs to meet AYP standards is likely to remain an ongoing focus of debate.

While few argue against having any participation rate requirement, it may be questioned whether it needs to be as high as 95%. In recent years, the overall percentage of enrolled pupils who attend public schools each day has been approximately 93.5%, and it is generally agreed that attendance rates are lower in schools serving relatively high proportions of disadvantaged pupils. Even though schools are explicitly allowed to administer assessments on make-up days following the primary date of test administration, and it is probable that more schools and LEAs will meet this requirement as they become more fully aware of its significance, it is likely to continue to be very difficult for many schools and LEAs to meet a 95% test participation requirement.

State Variations in Assessments and Proficiency Standards

As noted above, it is likely that state variations in the percentage of schools failing to meet AYP standards are based not only on underlying differences in achievement levels, as well as a variety of technical factors in state AYP provisions, but also on differences in the degree of rigor or challenge in state pupil performance standards and assessments. Particularly now that all states receiving Title I-A grants must also participate in state-level administration of NAEP tests in 4th and 8th grade reading and math every two years, this variation can be illustrated for all states by comparing the percentage of pupils scoring at the proficient level on NAEP versus state assessments.

Such a comparison was conducted by a private organization, Achieve, Inc., based on 8th grade reading and math assessments administered in the spring of 2003.⁵⁷ For a variety of reasons (e.g., several states did not administer standards-based assessments in reading or math to 8th grade pupils in 2003), the analysis excluded several states; 29 states were included in the comparison for reading, and 32 states for math. According to this analysis, the percentage of pupils statewide who score

⁵⁷ Center on Education Policy, *From the Capital to the Classroom, Year 2 of the No Child Left Behind Act* (Jan. 2004), p. 61.

at a proficient or higher level on state assessments, using state-specific pupil performance standards, was generally much higher than the percentage deemed to be at the proficient or higher level on the NAEP tests, and employing NAEP's pupil performance standards. Of the states considered, the percentage of pupils scoring at a proficient or higher level on the state assessment was lower than on NAEP (implying a more rigorous state standard) for five states⁵⁸ (out of 32) in math and only two states (out of 29) in reading. Further, among the majority of states where the percentage of pupils at the proficient level or above was found to be higher on state assessments than on NAEP, the relationship between the size of the two groups varied widely — in some cases only marginally higher on the state assessment, and in others the percentage at the proficient level was more than twice as high on the state assessment as on NAEP. While some portion of these differences in performance may result from differences in the motivation of pupils to perform well (and of teachers to encourage high performance) on NAEP versus state assessments, comparisons to NAEP results help to illuminate the variations in state proficiency standards. It is not yet clear whether such comparisons will significantly encourage greater consistency in those standards.

A second issue is whether some states might choose to lower their standards of “proficient” performance, in order to reduce the number of schools identified as failing to meet AYP and make it easier to meet the ultimate NCLBA goal of all pupils at the proficient or higher level within 12 years. In the affected states, this would increase the percentage of pupils deemed to be achieving at a “proficient” level, and reduce the number of schools failing to meet AYP standards.

While states are generally free to take such actions without jeopardizing their eligibility for Title I-A grants, since performance standards are ultimately state-determined and have always varied significantly, such actions have elicited public criticism from ED. In a policy letter dated October 22, 2002, the Secretary of Education stated that

Unfortunately, some states have lowered the bar of expectations to hide the low performance of their schools. And a few others are discussing how they can ratchet down their standards in order to remove schools from their lists of low performers. Sadly, a small number of persons have suggested reducing standards for defining “proficiency” in order to artificially present the facts. . . . Those who play semantic games or try to tinker with state numbers to lock out parents and the public, stand in the way of progress and reform. They are the enemies of equal justice and equal opportunity. They are apologists for failure.⁵⁹

⁵⁸ In two additional states, the percentages were essentially the same.

⁵⁹ See [<http://www.ed.gov/news/pressreleases/2002/10/10232002a.html>].